

A similarity-based implementation of the Schaake shuffle

Roman Schefzik

*Heidelberg Institute for Theoretical Studies
Schloss-Wolfsbrunnengasse 35, 69118 Heidelberg, Germany*

roman.schefzik@h-its.org

July 9, 2015

Abstract

Contemporary weather forecasts are typically based on ensemble prediction systems, which consist of multiple runs of numerical weather prediction models that vary with respect to the initial conditions and/or the parameterization of the atmosphere. Ensemble forecasts are frequently biased and show dispersion errors and thus need to be statistically postprocessed. However, current postprocessing approaches are often univariate and apply to a single weather quantity at a single location and for a single prediction horizon only, thereby failing to account for potentially crucial dependence structures. Non-parametric multivariate postprocessing methods based on empirical copulas, such as ensemble copula coupling or the Schaake shuffle, can address this shortcoming. A specific implementation of the Schaake shuffle, called the SimSchaake approach, is introduced. The SimSchaake method aggregates univariately postprocessed ensemble forecasts using dependence patterns from past observations. Specifically, the observations are taken from historical dates at which the ensemble forecasts resembled the current ensemble prediction with respect to a specific similarity criterion. The SimSchaake ensemble outperforms all reference ensembles in an application to ensemble forecasts for surface temperature from the European Centre for Medium-Range Weather Forecasts.

Keywords: empirical copula, ensemble copula coupling, probabilistic weather forecasting, Schaake shuffle, similarity criterion, statistical ensemble postprocessing

1 Introduction

Contemporary weather forecasts are typically constructed from ensemble prediction systems, which have run operationally since the early 1990s. An ensemble consists of multiple runs of numerical weather prediction models which vary with respect to the initial conditions and/or the parameterization of the atmosphere. Consequently, ensembles take account of the two major sources of uncertainty (Gneiting and Raftery, 2005; Leutbecher and Palmer, 2008; Palmer, 2002). Ensemble forecasts are frequently biased and show dispersion errors (Hamill and Colucci, 1997). Thus, they require statistical postprocessing to realize their full capability. During the last decade, several ensemble postprocessing methods have been developed. Examples are (variants of) the ensemble model output statistics (EMOS) (Gneiting et al., 2005, among others) approach, which is also known as non-homogeneous

regression, or Bayesian model averaging (BMA) (Raftery et al., 2005, among others). However, EMOS and BMA, as well as other postprocessing methods, only apply to a single weather variable at a single location and for a single prediction horizon. Thus, they fail to account for spatial, temporal or inter-variable dependence structures, which are crucial in many applications such as flood warning (Schaaake et al., 2010), winter road maintenance (Berrocal et al., 2010) or the handling of renewable energy sources (Pinson, 2013). In recent years, there has been a keen interest in the development of multivariate postprocessing methods being able to address this shortcoming, and a lot of effort has been invested to that end. For instance, variants and modifications of EMOS and BMA, respectively, that can handle spatial (Berrocal et al., 2007, 2008; Feldmann et al., 2015) or inter-variable (Baran and Möller, 2015; Möller et al., 2013; Schuhen et al., 2012; Sloughter et al., 2013) dependencies are available. Furthermore, there are methods to capture temporal dependencies of consecutive lead times in postprocessed predictive distributions (Pinson et al., 2009; Schoelzel and Hense, 2011). All these multivariate postprocessing methods are parametric and work well in rather low dimensions and in settings in which the involved correlation matrix can be taken to be highly structured. However, they model one type of dependence (spatial, inter-variable or temporal) only and appear to be inadequate when considering high-dimensional situations, in which no particular structure can be exploited. These issues can be addressed using non-parametric techniques based on the use of empirical copulas (Scheffzik, 2015b). Following Wilks (2014), an empirical copula (Deheuvels, 1979; Rüschendorf, 2009) can be interpreted as a dependence template induced by a specific discrete multivariate data set. It can be employed to transfer a particular dependence pattern to samples which are drawn independently from a collection of univariate marginal distributions (Wilks, 2014). For example, in the ensemble copula coupling (ECC) approach of Scheffzik et al. (2013), such an empirical copula is derived from the unprocessed ensemble forecast and then applied to samples from univariate postprocessed predictive distributions, which can be gained via the standard EMOS or BMA approaches. This is equivalent to ordering these samples according to the rank dependence structure of the raw ensemble, thereby capturing the spatial, temporal and inter-variable flow dependence (Scheffzik et al., 2013). Proceeding in a similar manner, the Schaaake shuffle (Clark et al., 2004) employs an ordering based on past observations from a historical data archive. Consequently, the corresponding empirical copula in the Schaaake shuffle is induced by an observational database rather than by an ensemble forecast. However, the standard Schaaake shuffle fails to condition the multivariate dependence pattern on current or predicted atmospheric states. To address this shortcoming, Clark et al. (2004, page 260) proposed to develop an extension thereof, driven by the idea

“to preferentially select dates from the historical record that resemble forecasted atmospheric conditions and use the spatial correlation structure from this subset of dates to reconstruct the spatial variability for a specific forecast.”

Inspired by this suggestion, a specific implementation of the Schaaake shuffle, referred to as the SimSchaaake approach, is introduced in this paper. Essentially, the SimSchaaake method proceeds like the Schaaake shuffle, but the observations determining the dependence structure are taken from historical dates at which the ensemble forecasts resembled the current ensemble prediction with respect to a specific similarity criterion.

The remainder of the paper is organized as follows. In Section 2, we first discuss the general setting of empirical copula-based ensemble postprocessing and review ECC and the Schaaake shuffle as reference

examples. Then, we develop the SimSchaake approach. In Section 3, this new method is evaluated and compared to the reference methods in a case study. The paper closes with a discussion in Section 4.

2 Empirical copula-based ensemble postprocessing methods

Copulas are valuable and established tools for the modeling of stochastic dependence (Joe, 2014; Nelsen, 2006). They have been successfully employed in numerous application areas. A copula is an L -variate cumulative distribution function (CDF) with standard uniform univariate marginal CDFs, where $L \in \mathbb{N}$, $L \geq 2$. As is manifested in the famous Sklar's theorem (Sklar, 1959), a copula C links a multivariate CDF H to its univariate marginal CDFs F_1, \dots, F_L via the decomposition

$$H(u_1, \dots, u_L) = C(F_1(u_1), \dots, F_L(u_L))$$

for $u_1, \dots, u_L \in \mathbb{R}$. In a multivariate postprocessing setting, the sought multivariate CDF H can thus be constructed by specifying both the univariate marginal CDFs F_1, \dots, F_L and the copula C modeling the dependence. The CDFs F_1, \dots, F_L can be obtained by common univariate postprocessing for each location, weather variable and look-ahead time individually, for instance performed via EMOS or BMA. For the choice of C , a prominent example is the Gaussian copula, which has been applied to a wide range of problems in climatology, meteorology and hydrology (Genest and Favre, 2007; Möller et al., 2013; Pinson, 2012; Schoelzel and Friederichs, 2008; Schuhen et al., 2012).

In this paper, we focus on the case in which F_1, \dots, F_L are the empirical CDFs given by samples from univariate postprocessed CDFs and C is taken to be an empirical copula (Deheuvels, 1979; Rüschendorf, 2009), E_N . According to Wilks (2014), an empirical copula can be considered a multivariate dependence template derived from a specific discrete data set. To describe this formally, let $\mathbf{z} := \{(z_1^1, \dots, z_N^1), \dots, (z_1^L, \dots, z_N^L)\}$ be a data set consisting of L tuples of size N with entries in \mathbb{R} . Moreover, let $\text{rank}(z_n^\ell)$ denote the rank of z_n^ℓ in $\{z_1^\ell, \dots, z_N^\ell\}$ for $n \in \{1, \dots, N\}$ and $\ell \in \{1, \dots, L\}$, assuming for simplicity that there are no ties. Then, the empirical copula E_N induced by the data set \mathbf{z} is given by

$$\begin{aligned} E_N \left(\frac{i_1}{N}, \dots, \frac{i_L}{N} \right) &:= \frac{1}{N} \sum_{n=1}^N \mathbb{1}_{\{\text{rank}(z_n^1) \leq i_1, \dots, \text{rank}(z_n^L) \leq i_L\}} \\ &= \frac{1}{N} \sum_{n=1}^N \prod_{\ell=1}^L \mathbb{1}_{\{\text{rank}(z_n^\ell) \leq i_\ell\}} \end{aligned} \quad (1)$$

for integers $0 \leq i_1, \dots, i_L \leq N$, with $\mathbb{1}_A$ denoting the indicator function whose value is 1 if the event A occurs, and zero otherwise.

In this section, we review ensemble copula coupling (ECC) (Scheffzik et al., 2013) and the Schaake shuffle (Clark et al., 2004) as reference methods within the general frame of empirical copula-based multivariate ensemble postprocessing (Scheffzik, 2015b). In addition, we develop the SimSchaake method as a specific implementation of the Schaake shuffle. While ECC and the Schaake shuffle have been used as a benchmark in several papers (Schaake et al., 2007; Scheuerer and Hamill, 2015; Voisin et al., 2011; Vrac and Friederichs, 2015; Wilks, 2014), the SimSchaake method is new.

To allow for a formal description of the approaches, let us first set some notation which will be used throughout the whole section. Let $i \in \{1, \dots, I\}$ be a weather variable, $j \in \{1, \dots, J\}$ a location and $k \in \{1, \dots, K\}$ a look-ahead time. For simplicity, let $\ell := (i, j, k)$ and $\ell^* := (i, j)$ denote the corresponding multi-indices, and let $L := I \times J \times K$ and $L^* := I \times J$, respectively. Moreover, let M denote the number of raw ensemble members, and N the desired number of members the postprocessed ensemble shall consist of.

In general, multivariate empirical copula-based ensemble methods to postprocess a raw ensemble forecast $\mathbf{x} := \{(x_1^1, \dots, x_M^1), \dots, (x_1^L, \dots, x_M^L)\}$ initialized at a specific date t_0 proceed according to the following scheme.

Scheme 1. (Empirical copula-based multivariate ensemble postprocessing)

1. To implement dependence structures, derive an empirical copula E_N from a suitable data set $\mathbf{z} := \{(z_1^1, \dots, z_N^1), \dots, (z_1^L, \dots, z_N^L)\}$ via (1).
Equivalently, derive the univariate order statistics $z_{(1)}^\ell \leq \dots \leq z_{(N)}^\ell$ for $\ell \in \{1, \dots, L\}$, inducing the permutation $\pi_\ell(n) := \text{rank}(z_n^\ell)$ for $n \in \{1, \dots, N\}$. If there are ties, let them be resolved at random.
2. For each margin ℓ , perform univariate postprocessing of the raw ensemble forecast $x_1^\ell, \dots, x_M^\ell$ based on past forecasts and observations as training data, and obtain a postprocessed predictive CDF F_ℓ in each case.
3. Draw a sample $\tilde{x}_1^\ell, \dots, \tilde{x}_N^\ell$ from each marginal CDF F_ℓ . This can be for instance and most conveniently done by taking equidistant quantiles of F_ℓ of the form

$$\tilde{x}_1^\ell := F_\ell^{-1}\left(\frac{1}{N+1}\right), \dots, \tilde{x}_N^\ell := F_\ell^{-1}\left(\frac{N}{N+1}\right). \quad (2)$$

4. Apply the empirical copula E_N to the samples from step 3.
Equivalently, arrange these samples with respect to the rank dependence structure of the chosen data set \mathbf{z} in step 1. Using the permutation π_ℓ , the final postprocessed ensemble $\hat{x}_1^\ell, \dots, \hat{x}_N^\ell$ for each margin ℓ is thus given by

$$\hat{x}_1^\ell := \tilde{x}_{(\pi_\ell(1))}^\ell, \dots, \hat{x}_N^\ell := \tilde{x}_{(\pi_\ell(N))}^\ell.$$

2.1 Reference methods

Now we review ECC (Scheffzik et al., 2013) and the Schaake shuffle (Clark et al., 2004) as empirical copula-based postprocessing techniques within the scheme described above.

2.1.1 Ensemble copula coupling (ECC)

In the ECC approach (Scheffzik et al., 2013), the data set specifying the dependence structure is given by the raw ensemble forecast, that is, we have $\mathbf{z} = \mathbf{x} = \{(x_1^1, \dots, x_M^1), \dots, (x_1^L, \dots, x_M^L)\}$ in Scheme 1. Consequently, the sample size in step 3 and hence the size of the final postprocessed ensemble in step 4 of Scheme 1 is restricted to equal that of the unprocessed ensemble, that is, $N = M$.

The ECC procedure operates under a perfect model assumption, implicitly assuming that the ensemble is capable to represent actual spatial, inter-variable and temporal dependence structures adequately. This may or may not be expected, but surely does not hold each and every day. Moreover, ECC only applies to ensembles whose members can be considered exchangeable, that is, statistically indistinguishable.

Schefzik et al. (2013) and Schefzik (2015a) show that ECC provides an overarching frame for seemingly unrelated approaches scattered in the literature such as those of Pinson (2012), Roulin and Vannitsem (2012), Flowerdew (2014) or Van Schaeybroeck and Vannitsem (2014), to name just a few. It has been used as a reference technique in the recent papers by Wilks (2014), Feldmann et al. (2015) and Scheuerer and Hamill (2015).

2.1.2 The Schaake shuffle

In contrast to ECC, the data set determining the dependence pattern in the Schaake shuffle (Clark et al., 2004) is not given by the raw ensemble forecasts, but by past observations

$\mathbf{y} := \{(y_1^1, \dots, y_N^1), \dots, (y_1^L, \dots, y_N^L)\}$ taken from N different dates of a historical archive. That is, observations from the same N dates are employed for all locations and weather variables throughout the procedure of the Schaake shuffle for a fixed forecast instance. Hence, we have $\mathbf{z} = \mathbf{y}$ in Scheme 1. In particular, N does not need to equal the raw ensemble size M .

In the original implementation of the Schaake shuffle by Clark et al. (2004), the corresponding N dates, from which the observations are taken, are chosen from all years in the historical record, except for the year of the forecast of interest, and lie within seven days before and after the verification date t , regardless of the year. A more general implementation of the Schaake shuffle may use observations from arbitrary (or randomly selected) past dates in the whole historical record. We will refer to this procedure as the Random Schaake method in the case study in Section 3.

The Schaake shuffle has been employed successfully in numerous applications (Robertson et al., 2013; Schaake et al., 2007; Verkade et al., 2013; Voisin et al., 2010, 2011; Vrac and Friederichs, 2015; Wilks, 2014).

2.2 The SimSchaake approach as a similarity-based implementation of the Schaake shuffle

As we have seen, the Schaake shuffle generates a postprocessed ensemble inheriting the rank dependence structure from historical observations. However, its standard implementations fail to condition the multivariate dependence pattern on current or predicted atmospheric states. Inspired by the quote of Clark et al. (2004) mentioned in the introductory Section 1, we address this challenge by linking the Schaake shuffle to similarity- or analog-based ensemble methods. In such approaches, one seeks ensemble forecasts in an archive of past data that are similar to the current one. The basic idea is that the realizing states of the atmosphere corresponding to such an analog ensemble can be assumed to be similar to the state to be predicted (Hamill and Whitaker, 2006). These techniques have become popular and important, as for instance witnessed by the papers of Bannayan and Hoogenboom (2008), Klausner et al. (2009), Hall et al. (2010), Delle Monache et al. (2011), Messner and Mayr (2011) and Delle Monache et al. (2013), and further by recent research (Alessandrini et al., 2015; Junk et al., 2015; Vanvyve et al., 2015). In this context, the question of the choice of appropriate similarity criteria in a

nearest neighbor sense arises, with the papers above providing some proposals.

The following approach, which will be referred to as the SimSchaake method in what follows, combines the idea of searching for similar ensembles and the Schaake shuffle. Like the Schaake shuffle and in contrast to standard ECC, it can be applied to any ensemble, regardless of whether it consists of exchangeable or non-exchangeable members, and the size N of the final postprocessed ensemble is not restricted to equal the raw ensemble size M . To describe the new SimSchaake approach formally in detail, let Λ be the length of the training period required for the univariate ensemble postprocessing, t_0 the initialization date of the ensemble forecast, and t the verification date. Let further D be the number of dates in the past of t_0 for which ensemble forecast and observation data is available. For the feasibility of the SimSchaake approach, it is required that ensemble forecast and observation data is available for at least $\max\{N, \Lambda\}$ dates in the past of t_0 , that is, $D \geq \max\{N, \Lambda\}$. Considering the prediction horizon to be fixed, the SimSchaake approach then proceeds according to the empirical copula-based postprocessing as in Scheme 1, where the data set \mathbf{z} in step 1 is derived as follows.

- i. For a fixed margin ℓ^* , let $\mathbf{x}^{\ell^*, \tau} := (x_1^{\ell^*, \tau}, \dots, x_M^{\ell^*, \tau})$ denote the (possibly standardized) M -member raw ensemble forecast valid on date τ . Let further

$$\mathbf{x}^\tau := (\mathbf{x}^{1, \tau}, \mathbf{x}^{2, \tau}, \dots, \mathbf{x}^{L^*, \tau}) = ((x_1^{1, \tau}, \dots, x_M^{1, \tau}), (x_1^{2, \tau}, \dots, x_M^{2, \tau}), \dots, (x_1^{L^*, \tau}, \dots, x_M^{L^*, \tau}))$$

denote the corresponding $(L^* \times M)$ -tuple consisting of the M -member ensemble forecasts of all L^* margins, that is, combinations of weather variable and location.

If weather variables with distinct units or magnitudes are involved, the components of $\mathbf{x}^{\ell^*, \tau}$ should be standardized for each multi-index ℓ^* .

- ii. For each date t_d in the past of the initialization date t_0 , where $d \in \{1, \dots, D\}$, compute a suitable fixed similarity criterion

$$\Delta^{t_d} : \mathbb{R}^{L^* \times M} \times \mathbb{R}^{L^* \times M} \rightarrow [0, \infty), \quad (\mathbf{x}^t, \mathbf{x}^{t_d}) \mapsto \Delta^{t_d}(\mathbf{x}^t, \mathbf{x}^{t_d}),$$

between the actual forecast \mathbf{x}^t valid on the verification date t and the forecast \mathbf{x}^{t_d} valid on the date t_d . The similarity criterion Δ^{t_d} is taken to be negatively oriented, that is, the lower the similarity criterion the more similar the ensemble forecasts. A similarity criterion value of exactly zero indicates that the ensemble forecasts are identical.

- iii. Choose those N dates $\tau_1, \dots, \tau_N \in \{t_1, \dots, t_D\}$ for which the data is most similar to that for the date t in the sense that the corresponding values of Δ^{τ_n} for $n \in \{1, \dots, N\}$ are the smallest among the values of Δ^{t_d} for $d \in \{1, \dots, D\}$.

Note that the information of all multi-indices ℓ^* simultaneously is employed to determine the dates τ_1, \dots, τ_N .

- iv. For each margin ℓ^* , let $y^{\ell^*, \tau_1}, \dots, y^{\ell^*, \tau_N}$ denote the corresponding N historical verifying observations valid on the dates τ_1, \dots, τ_N determined in step iii. For simplicity, write $y_n^{\ell^*} := y^{\ell^*, \tau_n}$ for $n \in \{1, \dots, N\}$ and build the data vector $\mathbf{y}^{\ell^*} := (y_1^{\ell^*}, \dots, y_N^{\ell^*})$ for the verification date t . The data set \mathbf{z} in step 1 from Scheme 1 is then obtained by aggregating the historical observation databases of all margins ℓ^* , that is, $\mathbf{z} = (\mathbf{y}^1, \dots, \mathbf{y}^{L^*})$. From this template \mathbf{z} , the empirical copula related to the SimSchaake approach is derived via (1).

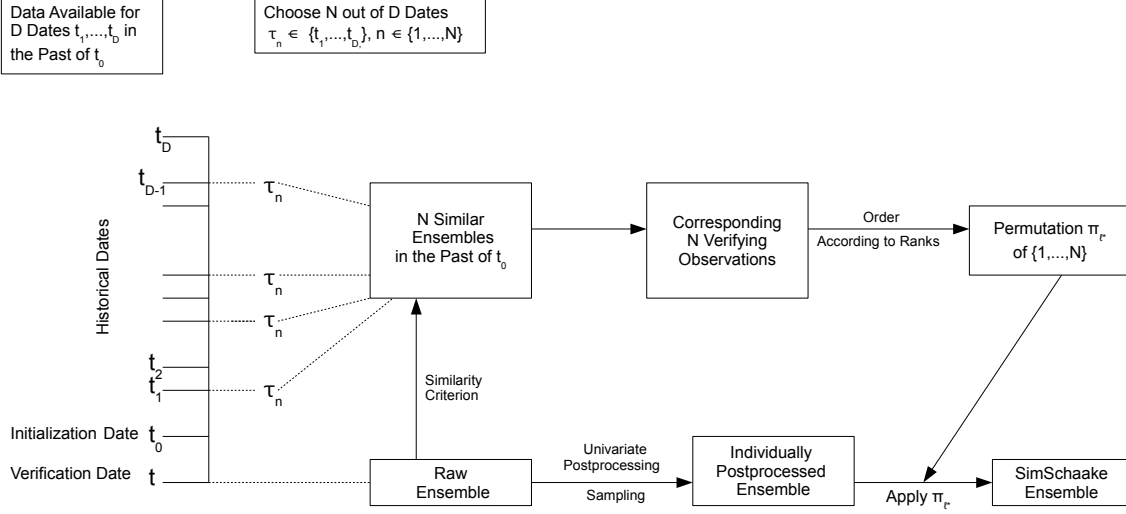


Figure 1: Scheme of the SimSchaake approach

Having obtained the data set \mathbf{z} and the respective empirical copula according to the above procedure, steps 2 to 4 from Scheme 1 are performed in order to generate the final postprocessed SimSchaake ensemble. A scheme of the SimSchaake approach is given in Fig. 1.

An appropriate choice of the similarity criterion Δ^{t_d} in step ii, which is then consistently used throughout the whole SimSchaake approach, is crucial. We here consider

$$\Delta^{t_d}(\mathbf{x}^t, \mathbf{x}^{t_d}) := \sqrt{\frac{1}{L^*} \sum_{\ell^*=1}^{L^*} (\bar{x}^{\ell^*,t} - \bar{x}^{\ell^*,t_d})^2 + \frac{1}{L^*} \sum_{\ell^*=1}^{L^*} (s^{\ell^*,t} - s^{\ell^*,t_d})^2}, \quad (3)$$

where

$$\bar{x}^{\ell^*,\tau} := \frac{1}{M} \sum_{m=1}^M x_m^{\ell^*,\tau} \quad \text{and} \quad s^{\ell^*,\tau} := \sqrt{\frac{1}{M-1} \sum_{m=1}^M (x_m^{\ell^*,\tau} - \bar{x}^{\ell^*,\tau})^2}$$

denote the empirical mean and standard deviation, respectively, of the ensemble forecast $\mathbf{x}^{\ell^*,\tau}$ for the fixed multi-index ℓ^* at date τ . As it does not depend on how the ensemble members are labeled, the similarity criterion in (3) can be applied to ensembles consisting of exchangeable members. In particular, it is suitable for the temperature predictions from the European Centre for Medium-Range Weather Forecasts (ECMWF) used in the case study in the next section. While the use of the empirical mean to some degree accounts for seasonal aspects when considering temperature forecasts only, the empirical standard deviations reflect the uncertainties within the ensemble forecasts. Thus, these issues are addressed by (3) when comparing two ensemble forecasts. Alternative proposals for similarity criteria, also for the case of ensembles with non-exchangeable members, can be found in the references mentioned before.

It is possible to transform the values of a similarity criterion Δ^{t_d} from $[0, \infty)$ to $(0, 1]$ by employing the standardization $\hat{\Delta}^{t_d} := \exp(-\Delta^{t_d})$. With respect to $\hat{\Delta}^{t_d}$, similarity values near to 1 indicate a very

high similarity between \mathbf{x}^t and \mathbf{x}^{t_d} , while similarity values near to 0 point at no similarity. Accordingly, if using $\tilde{\Delta}^{t_d}$, we then have to choose the dates τ_1, \dots, τ_N corresponding to the highest, and not to the lowest, values of $\tilde{\Delta}^{t_d}$ in step iii.

As mentioned, the SimSchaake approach addresses two shortcomings in the standard ECC method. First, it can also be applied to ensembles consisting of non-exchangeable members, as the reordering is not based on the ensemble forecasts, but on observations. Second, with the SimSchaake technique we can principally create ensembles of arbitrary size, as long as there are sufficiently many historical observations in the past. In contrast to ECC, the postprocessed ensemble is thus not restricted to have the same number of members as the raw ensemble.

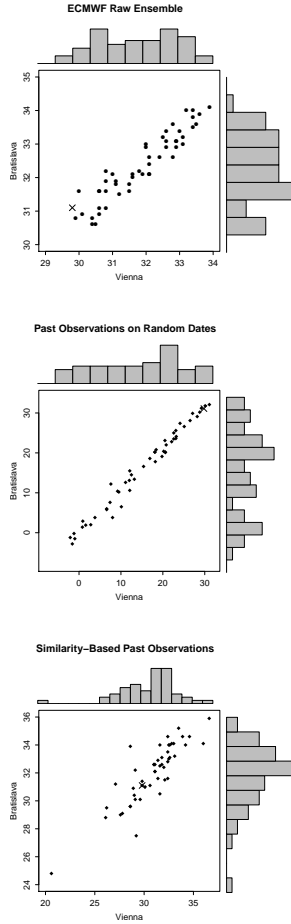
In Fig. 2, an illustration of the three empirical copula-based postprocessing methods presented in this section – ECC in the first row, the Schaake shuffle in the second row and the SimSchaake method in the third row – is given. We consider 24 hour ahead surface temperature forecasts (in degrees Celsius; °C) at Vienna (Austria) and Bratislava (Slovakia) valid on the hot summer day of 9 July 2011 at 1200 UTC. In the subfigures, the ensemble forecast is indicated by the dots, the verifying observation by the cross, and past observations from a historical database by the tetragons. As the ECMWF raw ensemble used here has size $M = 50$, the illustrations are for convenience based on $N = M = 50$ -member ensembles for both ECC and the Schaake shuffle-based methods. The first column of Fig. 2 shows the corresponding database that is used to determine the dependence structure and the empirical copula, respectively, of the postprocessing approach: the ECMWF raw ensemble in the case of ECC, past observations on random dates for the Random Schaake method, and past observations on specific dates selected according to the similarity criterion in (3) in the case of the SimSchaake approach. The second column shows three times the same individually postprocessed ensemble forecast. This is generated by randomly pairing the equidistant quantiles (2) from the predictive CDFs obtained by univariate EMOS postprocessing (Gneiting et al., 2005) at Vienna and Bratislava separately. Such an ensemble forecast does not take account of the pronounced positive spatial correlation structure. In the third column, the final postprocessed ensemble forecasts are shown, obtained by applying the corresponding empirical copula to the samples derived from the individual EMOS postprocessing. These empirical copula-based postprocessed ensembles have the same marginal distributions as the individual EMOS postprocessed ensemble, as witnessed by the respective histograms at the top and to the right of each subfigure. Additionally, they exhibit the same spatial correlation pattern as the underlying database specified in the first column, thus respecting dependencies.

3 Case study

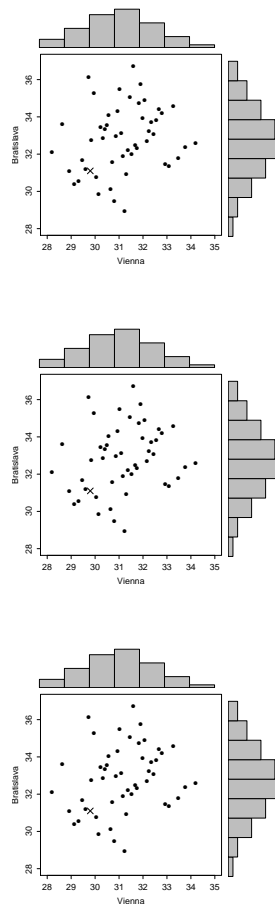
3.1 Setting

In our case study, we employ predictions of the European Centre for Medium-Range Weather Forecasts (ECMWF) core ensemble (ECMWF Directorate, 2012), whose $M = 50$ members can be considered exchangeable. We focus on 24 hour-ahead temperature forecasts jointly at Vienna (Austria), Bratislava (Slovakia) and Budapest (Hungary), and consequently on spatial dependencies only. For each location, the ground truth is given by the corresponding surface synoptic observations (SYNOP). The approximate distance from Vienna to Bratislava is 50 kilometers, from Bratislava to Budapest 170 kilometers, and from Vienna to Budapest 210 kilometers. There are pronounced positive pairwise correlations be-

(a) Database for Empirical Copula (Dependence Model)



(b) Individually EMOS Postprocessed Ensemble



(c) EMOS and Empirical Copula-Based Postprocessed Ensemble

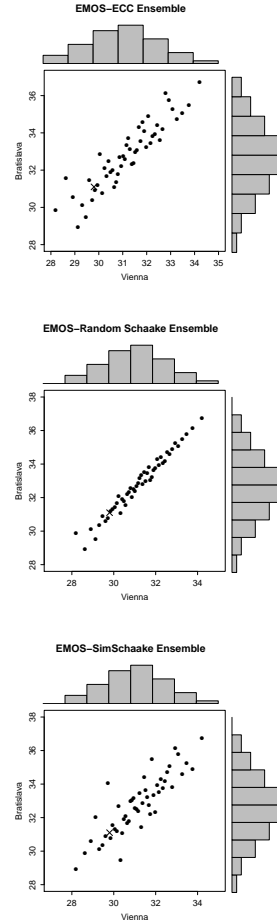


Figure 2: Illustration of ECC (first row), the Random Schaake method (second row) and the Sim-Schaake method (third row) using 24 hour ahead temperature forecasts (in $^{\circ}\text{C}$) at Vienna and Bratislava valid on 9 July 2011, 1200 UTC. Ensemble forecasts are indicated by the dots, and the verifying observation by the cross. The tetragons symbolize past observations from a historical database. At the top and to the right of each subfigure, the corresponding marginal histograms are shown.

tween the observations at the three different locations, which are the stronger the closer the respective stations are. These correlation patterns are basically reflected well in the ensemble forecasts.

We consider those 3985 test days during the period from 1 January 2003 to 31 December 2013 for which all required forecast and observation data is available at all the three stations. In our case study, all data is valid on 1200 UTC. Univariate postprocessing is performed via EMOS (Gneiting et al., 2005) using the R package `ensembleMOS` (R Core Team, 2014; Yuen et al., 2013), employing a rolling window consisting of the last $\Lambda = 50$ days before the verification day as training period. For each marginal EMOS postprocessed predictive CDF F_ℓ , we follow the quantization (2) and take the N equidistant $(n/(N+1))$ -quantiles, where $n = 1, \dots, N$, as samples, focusing on the cases of $N = M = 50$ and $N = 80$, respectively, here.

For these desired ensemble sizes, we assess and compare the predictive performances of

- the ECMWF raw ensemble,
- the Individual EMOS ensemble, which assumes independence,
- the EMOS-ECC ensemble, which assumes a dependence structure according to that in the raw ensemble,
- the EMOS-Random Schaake ensemble, which assumes a dependence structure according to randomly selected historical observation data, and
- the EMOS-SimSchaake ensemble, which assumes a dependence structure according to specific historical observation data valid on dates for which the ensemble forecast resembled the current one with respect to similarity criterion (3).

Obviously, results for the raw and the EMOS-ECC ensemble can only be reported for the case of $N = M = 50$, whereas the other ensembles are additionally evaluated for the final ensemble size of $N = 80$. For the two approaches employing the Schaake shuffle notion, the past dates from which the corresponding verifying observations are taken, are searched for among all available historical data, where ensemble forecast and observation data is available from 1 January 2002 to 31 December 2013. Hence, the database used for the Schaake shuffle-based methods grows over time. Recall that the EMOS-Random Schaake method just randomly selects those past dates, whereas the EMOS-SimSchaake approach chooses them based on the ensemble similarity criterion (3).

3.2 Evaluation tools

A probabilistic forecast distribution or an ensemble forecast, respectively, should be as sharp as possible, subject to calibration, which refers to statistical consistency between the forecasts and the observation (Gneiting et al., 2007). To assess the predictive performances of our different ensembles, several verification tools are available (Wilks, 2011).

In univariate settings, calibration can be checked via the verification rank histogram (Anderson, 1996; Hamill, 2001; Talagrand et al., 1997). As we focus on the evaluation of multivariate quantities in this paper, ensemble calibration in our case study is checked via the multivariate (Gneiting et al., 2008), band depth and average rank histograms (Thorarinsdottir et al., 2015). When an ensemble forecast is calibrated, the multivariate, band depth or average rank, respectively, is uniformly distributed.

Calibration can thus be diagnosed by compositing over forecast cases, plotting the corresponding multivariate, band depth or average rank histogram, respectively, and checking for deviations from uniformity, that is, flatness of the histogram. For an interpretation of the different shapes a rank histogram for multivariate quantities can exhibit, see Gneiting et al. (2008) and Thorarinsdottir et al. (2015).

The overall forecast skill can be assessed via proper scoring rules (Gneiting and Raftery, 2007), which are able to assess calibration and sharpness simultaneously and are taken to be negatively oriented here, that is, the lower the score the better the predictive performance. A widely used proper scoring rule for univariate quantities is the continuous ranked probability score (CRPS) (Gneiting and Raftery, 2007; Matheson and Winkler, 1976).

In this paper, we employ the energy score (ES) (Gneiting et al., 2008), which is the analog of the CRPS for multivariate quantities. For an N -member ensemble forecast

$$\mathbf{x}_1 := (x_1^1, \dots, x_1^L), \dots, \mathbf{x}_N := (x_N^1, \dots, x_N^L) \in \mathbb{R}^L$$

and an observation

$$\mathbf{y} := (y_1, \dots, y_L) \in \mathbb{R}^L,$$

the ES is computed as

$$\text{ES} = \frac{1}{N} \sum_{n=1}^N \|\mathbf{x}_n - \mathbf{y}\| - \frac{1}{2N^2} \sum_{\nu=1}^N \sum_{n=1}^N \|\mathbf{x}_\nu - \mathbf{x}_n\|,$$

with $\|\cdot\|$ denoting the Euclidean norm. As the ES reveals weaknesses in detecting misspecifications in the correlation structure (Pinson and Tastu, 2013; Scheuerer and Hamill, 2015), we additionally consider the variogram score (VS) (Scheuerer and Hamill, 2015) to address this, which is given by

$$\text{VS} = \sum_{\ell=1}^L \sum_{\lambda=1}^L w_{\ell\lambda} \left(|y_\ell - y_\lambda|^{\frac{1}{2}} - \frac{1}{N} \sum_{n=1}^N |x_n^\ell - x_n^\lambda|^{\frac{1}{2}} \right)^2,$$

where the $w_{\ell\lambda}$'s are (optional) non-negative weights. For the case study in this paper, in which we focus on spatial dependencies, we follow the suggestion of Scheuerer and Hamill (2015) and let the weights be proportional to the inverse spatial distances between the corresponding locations. That is, we choose

$$w_{\ell\lambda} := \frac{\frac{1}{\text{dist}(\ell, \lambda)}}{\sum_{\substack{\ell, \lambda=1 \\ \ell \neq \lambda}}^L \frac{1}{\text{dist}(\ell, \lambda)}},$$

for $\ell \neq \lambda$ and $w_{\ell\lambda} := 0$ for $\ell = \lambda$, with $\text{dist}(\ell, \lambda)$ denoting the spatial distance between location ℓ and location λ , where all distances have to be measured in the same unit. For the specific implementation in our $L = 3$ -dimensional setting here, we employ the distances between Vienna, Bratislava and Budapest as mentioned in the preceding subsection.

In our case study, average scores over all forecast cases within our specific test period are reported.

Table 1: Average energy scores (ES) and variogram scores (VS) for 24 hour ahead temperature forecasts at Vienna, Bratislava and Budapest jointly over 3985 test days during the period from 1 January 2003 to 31 December 2013. The scores for the Individual EMOS and the EMOS-Random Schaake ensembles are averaged over 100 runs.

		ES	VS
$M = 50$	ECMWF Raw Ensemble	2.241	0.333
$N = M = 50$	Individual EMOS Ensemble	1.976	0.323
	EMOS-ECC Ensemble	1.957	0.270
	EMOS-Random Schaake Ensemble	1.998	0.300
	EMOS-SimSchaake Ensemble	1.952	0.265
$N = 80$	Individual EMOS Ensemble	1.971	0.327
	EMOS-Random Schaake Ensemble	1.996	0.300
	EMOS-SimSchaake Ensemble	1.947	0.266

3.3 Results

As we focus on the multivariate setting in this paper, we do not explicitly show the results for the univariate EMOS postprocessing at the three stations individually here. In a nutshell, the EMOS postprocessed ensemble forecasts exhibit a better predictive performance than the unprocessed raw ensemble predictions, in that they are better calibrated and have smaller CRPS values.

In Table 1, the average energy scores (ES) and variogram scores (VS) as overall performance measures are shown. The results for the Individual EMOS and the EMOS-Random Schaake ensemble are averaged over 100 runs for each forecast instance, in order to account for randomizations. Precisely, for the Individual EMOS ensemble, the results for 100 different aggregations (that is, assignments of the member indices) of the equidistant quantiles obtained by the univariate EMOS postprocessing are averaged. In case of the EMOS-Random Schaake ensemble, the average is taken over 100 different selections of the random historical dates that are used to define the observation-based dependence model. Calibration is assessed via the multivariate, band depth and average rank histograms, respectively, in Figs. 3, 4 and 5.

Both in the case of $N = M = 50$ and $N = 80$, all postprocessing methods outperform the raw ensemble in terms of the scores. The raw ensemble is clearly uncalibrated, more precisely underdispersed, as witnessed by the U-shaped multivariate and average rank histograms and the skewed band depth rank histogram with an overpopulation of the lowest ranks (Gneiting et al., 2008; Thorarinsdottir et al., 2015).

Considering the results for the postprocessed ensembles consisting of $N = M = 50$ members first, the Individual EMOS ensemble is uncalibrated, yielding U-shaped rank histograms. In the case of the band depth and average rank histograms, respectively, this points at an underestimation of the correlation structure (Thorarinsdottir et al., 2015), which is plausible, as this approach does not account for dependencies. With respect to the VS, the Individual EMOS ensemble performs worse than the

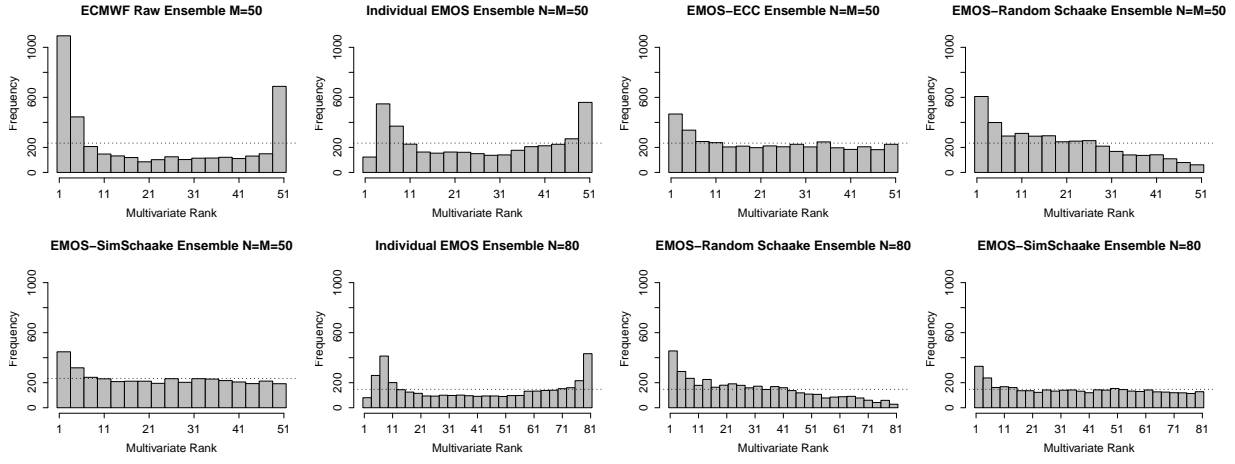


Figure 3: Multivariate rank histograms for 24 hour ahead temperature forecasts at Vienna, Bratislava and Budapest jointly over 3985 test days during the period from 1 January 2003 to 31 December 2013.

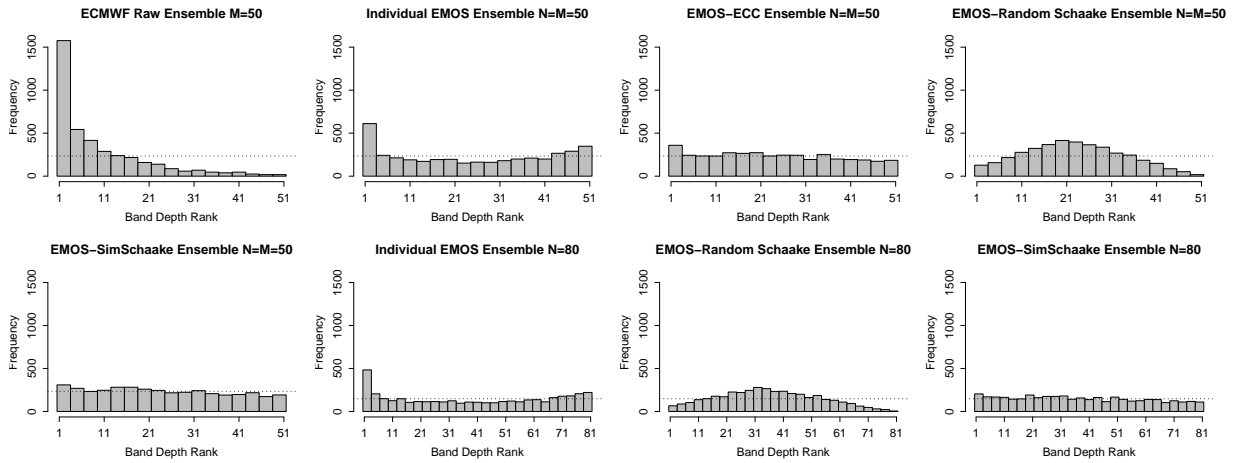


Figure 4: Band depth rank histograms for 24 hour ahead temperature forecasts at Vienna, Bratislava and Budapest jointly over 3985 test days during the period from 1 January 2003 to 31 December 2013

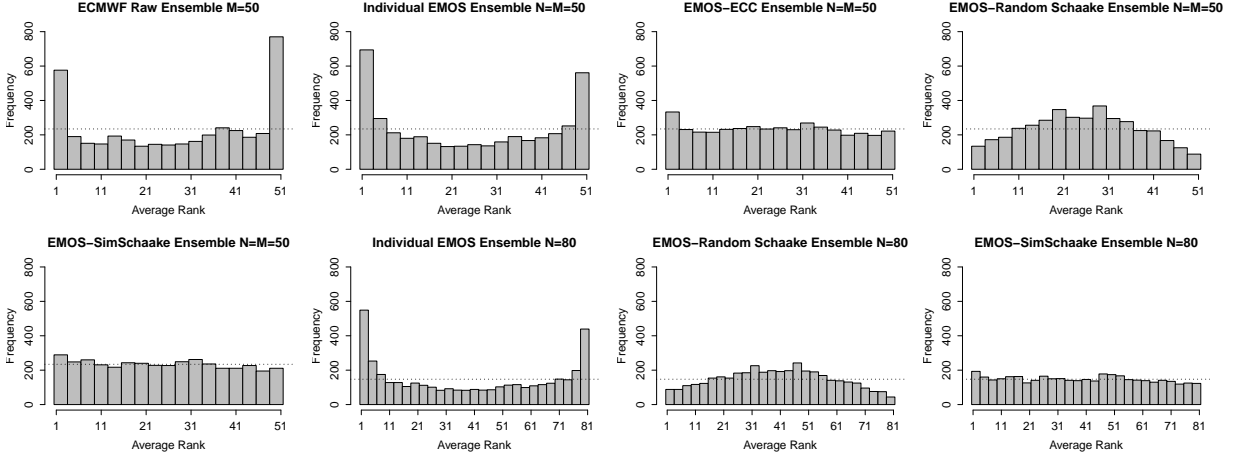


Figure 5: Average rank histograms for 24 hour ahead temperature forecasts at Vienna, Bratislava and Budapest jointly over 3985 test days during the period from 1 January 2003 to 31 December 2013

other postprocessed ensembles, which assume specific correlation structures. In terms of the ES, the EMOS-ECC and the EMOS-SimSchaake ensembles outperform the Individual EMOS ensemble, while the distinctions are less pronounced than for the VS. This may be due to the discrimination inability of the ES with respect to correlations between different locations (Pinson and Tastu, 2013; Scheuerer and Hamill, 2015), as hinted at in the preceding subsection.

Comparing the three empirical copula-based postprocessing methods taking account of dependence patterns, the EMOS-ECC ensemble is well calibrated, apart from a slight overpopulation of the lowest ranks in all rank histograms. In contrast, the calibration of the EMOS-Random Schaake ensemble is not that good, as witnessed by the inverse U-shaped band depth and average rank histograms, indicating an overestimation of the correlation structures (Thorarinsdottir et al., 2015). Moreover, the multivariate rank histogram of the EMOS-Random Schaake ensemble is skewed with too many low ranks. The EMOS-SimSchaake ensemble is calibrated best, with band depth and average rank histograms being close to uniform and an essentially flat multivariate rank histogram with a slight overpopulation of the lowest ranks only. The ranking of the three ensembles allowing for dependencies in terms of calibration is also reflected in the scores. Both for the ES and the VS, the EMOS-SimSchaake ensemble performs best, followed by the EMOS-ECC ensemble and finally the EMOS-Random Schaake ensemble.

The results and conclusions on calibration described above continue to hold analogously for the extended Individual EMOS, EMOS-Random Schaake and EMOS-SimSchaake ensembles comprising $N = 80$ members. Similarly, the ranking of the predictive skill of these three extended postprocessed ensembles remains unchanged with respect to the ES and VS, respectively, compared to the case of $N = M = 50$, with the EMOS-SimSchaake ensemble still performing by far best. The ES and VS values of the $N = 80$ -member ensembles are very similar to those of their counterparts in the case of $N = M = 50$. Although an extension of the ensemble size is generally useful, there is no pronounced need in our case here to increase the ensemble size of $N = M = 50$, which appears to be already reasonably large.

In a nutshell, the EMOS-SimSchaake ensemble based on the new method introduced in this paper

performs best among all reference ensembles, both with respect to calibration and in terms of scores. In particular, the EMOS-SimSchaake ensemble outperforms the EMOS-Random Schaake ensemble. Hence, there appears to be a clear benefit of using the specific past dates on which the ensemble forecasts resembled the current one to create the historical observation database modeling the dependencies, rather than picking these dates randomly. The EMOS-SimSchaake ensemble also outperforms the EMOS-ECC ensemble.

4 Discussion

We have discussed and compared empirical copula-based ensemble postprocessing methods that are able to account for dependencies. While ECC and the Schaake shuffle have been reviewed in a general frame, the SimSchaake scheme has been newly developed in this paper as a multivariate postprocessing tool. Essentially, the SimSchaake procedure aggregates samples from univariate postprocessed distributions, where the underlying dependence structure and the involved empirical copula, respectively, are derived from historical observations at dates in the past which showed a similar ensemble forecast to the current one. In our case study, the SimSchaake ensemble has performed best overall and better than the ECC ensemble, while having the benefit of a broader applicability, in that it can also be employed on ensembles comprising non-exchangeable members and is not restricted to have the same size as the raw ensemble.

The SimSchaake method depends on the design of a suitable similarity criterion, where the choice of (3) has proven to be useful, yielding good results. However, the predictive performance of the SimSchaake ensemble might be improved by using a more sophisticatedly designed similarity criterion, perhaps including a suitable weighting function (or monotone transformations thereof) that accounts for seasonal aspects. Moreover, the similarity criterion in (3) is tailored to ensembles consisting of exchangeable members such as the ECMWF ensemble in our case study. For ensembles comprising non-exchangeable members, other criteria might provide more reasonable and better choices.

As mentioned, a drawback of the standard ECC postprocessed ensemble employed here is that it is constrained to have the same size as the unprocessed ensemble, while the Schaake shuffle and the SimSchaake ensembles are not. However, there have been first attempts to design ECC-like ensembles having more members than the raw ensemble (Schefzik, 2015a; Wilks, 2014). Similar to the work in Wilks (2014), an issue for future work may be to design and conduct a further case study including these approaches to allow for a broader comparison of ECC- and Schaake shuffle-based concepts.

Acknowledgments

I gratefully acknowledge support by VolkswagenStiftung under the project “Mesoscale Weather Extremes: Theory, Spatial Modeling and Prediction”. Initial work on this paper was done during my time as Ph.D. student at Heidelberg University, funded by Deutsche Forschungsgemeinschaft through Research Training Group (RTG) 1953. I thank Tilmann Gneiting, Stephan Hemri, Sebastian Lerch and Martin Leutbecher for valuable comments, suggestions and discussions. The forecast data used in the case study have been made available by the European Centre for Medium-Range Weather Forecasts. I thank Stephan Hemri for help with the data and Michael Scheuerer for R code for the rank histograms for multivariate quantities.

References

- S. Alessandrini, L. Delle Monache, S. Sperati, and J. N. Nissen. A novel application of an analog ensemble for short-term wind power forecasting. *Renew. Energ.*, 76:768–781, 2015.
- J. L. Anderson. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, 9:1518–1530, 1996.
- M. Bannayan and G. Hoogenboom. Predicting realizations of daily weather data for climate forecasts using the non-parametric nearest-neighbour re-sampling technique. *Int. J. Climatol.*, 28:1357–1368, 2008.
- S. Baran and A. Möller. Joint probabilistic forecasting of wind speed and temperature using Bayesian model averaging. *Environmetrics*, 26:120–132, 2015.
- V. J. Berrocal, A. E. Raftery, and T. Gneiting. Combining spatial statistical and ensemble information in probabilistic weather forecasts. *Mon. Wea. Rev.*, 135:1386–1402, 2007.
- V. J. Berrocal, A. E. Raftery, and T. Gneiting. Probabilistic quantitative precipitation field forecasting using a two-stage spatial model. *Ann. Appl. Stat.*, 2:1170–1193, 2008.
- V. J. Berrocal, A. E. Raftery, T. Gneiting, and R. C. Steed. Probabilistic weather forecasting for winter road maintenance. *J. Amer. Stat. Assoc.*, 105:522–537, 2010.
- M. P. Clark, S. Gangopadhyay, L. E. Hay, B. Rajagopalan, and R. L. Wilby. The Schaake shuffle: A method for reconstructing space-time variability in forecasted precipitation and temperature fields. *J. Hydrometeor.*, 5:243–262, 2004.
- P. Deheuvels. La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d’indépendance. *Acad. Roy. Belg. Bull. Cl. Sci. (5)*, 65:274–292, 1979.
- L. Delle Monache, T. Nipen, Y. Liu, G. Roux, and R. Stull. Kalman filter and analog schemes to postprocess numerical weather predictions. *Mon. Wea. Rev.*, 139:3554–3570, 2011.
- L. Delle Monache, F. A. Eckel, D. L. Rife, B. Nagarajan, and K. Searight. Probabilistic weather prediction with an analog ensemble. *Mon. Wea. Rev.*, 141:3498–3516, 2013.
- ECMWF Directorate. Describing ECMWF’s forecasts and forecasting system. *ECMWF Newsletter*, 133:11–13, 2012.
- K. Feldmann, M. Scheuerer, and T. L. Thorarinsdottir. Spatial postprocessing of ensemble forecasts for temperature using nonhomogeneous Gaussian regression. *Mon. Wea. Rev.*, 143:955–971, 2015.
- J. Flowerdew. Calibrating ensemble reliability whilst preserving spatial structure. *Tellus A*, 66:22662, 2014.
- C. Genest and A.-C. Favre. Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Eng.*, 12:347–368, 2007.

- T. Gneiting and A. E. Raftery. Weather forecasting with ensemble methods. *Science*, 310:248–249, 2005.
- T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *J. Amer. Stat. Assoc.*, 102:359–378, 2007.
- T. Gneiting, A. E. Raftery, A. H. Westveld, and T. Goldman. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, 133:1098–1118, 2005.
- T. Gneiting, F. Balabdaoui, and A. E. Raftery. Probabilistic forecasts, calibration and sharpness. *J. Roy. Stat. Soc. Ser. B*, 69:243–268, 2007.
- T. Gneiting, L. I. Stanberry, E. P. Grimit, L. Held, and N. A. Johnson. Assessing probabilistic forecasts of multivariate quantities, with applications to ensemble predictions of surface winds (with discussion and rejoinder). *Test*, 17:211–264, 2008.
- T. J. Hall, R. N. Thessin, G. J. Bloy, and C. N. Mutchler. Analog sky condition forecasting based on a k -nn algorithm. *Wea. Forecasting*, 25:1463–1478, 2010.
- T. M. Hamill. Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, 129:550–560, 2001.
- T. M. Hamill and S. J. Colucci. Verification of Eta-RSM short-range ensemble forecasts. *Mon. Wea. Rev.*, 125:1312–1327, 1997.
- T. M. Hamill and J. S. Whitaker. Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. *Mon. Wea. Rev.*, 134:3209–3229, 2006.
- H. Joe. *Dependence Modeling with Copulas*. CRC Press, 2014.
- C. Junk, L. Delle Monache, and S. Alessandrini. Analog-based Ensemble Model Output Statistics. *Mon. Wea. Rev.*, 2015. doi: 10.1175/MWR-D-15-0095.1. in press.
- Z. Klausner, H. Kaplan, and E. Fattal. The similar days method for predicting near surface wind vectors. *Meteor. Appl.*, 16:569–579, 2009.
- M. Leutbecher and T. N. Palmer. Ensemble forecasting. *J. Comput. Phys.*, 227:3515–3539, 2008.
- J. E. Matheson and R. L. Winkler. Scoring rules for continuous probability distributions. *Manag. Sci.*, 22:1087–1096, 1976.
- J. W. Messner and G. J. Mayr. Probabilistic forecasts using analogs in the idealized Lorenz96 setting. *Mon. Wea. Rev.*, 139:1960–1971, 2011.
- A. Möller, A. Lenkoski, and T. L. Thorarinsdottir. Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Quart. J. Roy. Meteor. Soc.*, 139:982–991, 2013.
- R. B. Nelsen. *An Introduction to Copulas*. Springer, New York, 2nd edition, 2006.

- T. N. Palmer. The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quart. J. Roy. Meteor. Soc.*, 128:747–774, 2002.
- P. Pinson. Adaptive calibration of (u, v) -wind ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, 138:1273–1284, 2012.
- P. Pinson. Wind energy: Forecasting challenges for its operational management. *Stat. Sci.*, 28:564–585, 2013.
- P. Pinson and J. Tastu. Discrimination ability of the Energy score. Technical report, Technical University of Denmark, 2013.
- P. Pinson, H. Madsen, H. A. Nielsen, G. Papaefthymiou, and B. Klöckl. From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, 12:51–62, 2009.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL <http://www.R-project.org>.
- A. E. Raftery, T. Gneiting, F. Balabdaoui, and M. Polakowski. Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, 133:1155–1174, 2005.
- D. E. Robertson, D. L. Shrestha, and Q. J. Wang. Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting. *Hydrol. Earth Syst. Sci.*, 17:3587–3603, 2013.
- E. Roulin and S. Vannitsem. Postprocessing of ensemble precipitation predictions with extended logistic regression based on hindcasts. *Mon. Wea. Rev.*, 140:874–888, 2012.
- L. Rüschendorf. On the distributional transform, Sklar’s theorem, and the empirical copula process. *J. Statist. Plann. Inference*, 139:3921–3927, 2009.
- J. C. Schaake, J. Demargne, R. Hartman, M. Mullusky, E. Welles, L. Wu, H. Herr, X. Fan, and D. J. Seo. Precipitation and temperature ensemble forecasts from single-valued forecasts. *Hydrol. Earth Syst. Sci. Discuss.*, 4:655–717, 2007.
- J. C. Schaake, J. Pailleux, J. Thielen, R. Arritt, T. M. Hamill, L. Luo, E. Martin, D. McCollor, and F. Pappenberger. Summary of recommendations of the first Workshop on Postprocessing and Downscaling Atmospheric Forecasts for Hydrologic Applications held at Météo-France, Toulouse, France, 15–18 June 2009. *Atmos. Sci. Lett.*, 11:59–63, 2010.
- R. Schefzik. *Physically Coherent Probabilistic Weather Forecasts Using Multivariate Discrete Copula-Based Ensemble Postprocessing Methods*. PhD thesis, Heidelberg University, 2015a.
- R. Schefzik. Multivariate discrete copulas, with applications in probabilistic weather forecasting. *Ann. I.S.U.P.*, 2015b. in press.
- R. Schefzik, T. L. Thorarinsdottir, and T. Gneiting. Uncertainty quantification in complex simulation models using ensemble copula coupling. *Stat. Sci.*, 28:616–640, 2013.

- M. Scheuerer and T. M. Hamill. Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Mon. Wea. Rev.*, 143:1321–1334, 2015.
- C. Schoelzel and P. Friederichs. Multivariate non-normally distributed random variables in climate research — Introduction to the copula approach. *Nonlinear Proc. Geoph.*, 15:761–772, 2008.
- C. Schoelzel and A. Hense. Probabilistic assessment of regional climate change in Southwest Germany by ensemble dressing. *Climate Dyn.*, 36:2003–2014, 2011.
- N. Schuhen, T. L. Thorarinsdottir, and T. Gneiting. Ensemble model output statistics for wind vectors. *Mon. Wea. Rev.*, 140:3204–3219, 2012.
- A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris*, 8: 229–231, 1959.
- J. M. Sloughter, T. Gneiting, and A. E. Raftery. Probabilistic wind vector forecasting using ensembles and Bayesian model averaging. *Mon. Wea. Rev.*, 141:2107–2119, 2013.
- O. Talagrand, R. Vautard, and B. Strauss. Evaluation of probabilistic prediction systems. In *Proceeding of ECMWF Workshop on Predictability*, pages 1–25. Reading, UK, European Centre for Medium-Range Weather Forecasts, 1997.
- T. L. Thorarinsdottir, M. Scheuerer, and C. Heinz. Assessing the calibration of high-dimensional ensemble forecasts using rank histograms. *J. Comput. Graph. Statist.*, 2015. in press.
- B. Van Schaeybroeck and S. Vannitsem. Ensemble post-processing using member-by-member approaches: theoretical aspects. *Quart. J. Roy. Meteor. Soc.*, 2014. doi: 10.1002/qj.2397. in press.
- E. Vanvyve, L. Delle Monache, A. J. Monaghan, and J. O. Pinto. Wind resource estimates with an analog ensemble approach. *Renew. Energ.*, 74:761–773, 2015.
- J. S. Verkade, J. D. Brown, P. Reggiani, and A. H. Weerts. Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *J. Hydrol.*, 501:73–91, 2013.
- N. Voisin, J. C. Schaake, and D. P. Lettenmeier. Calibration and downscaling methods for quantitative ensemble precipitation forecasts. *Wea. Forecasting*, 25:1603–1627, 2010.
- N. Voisin, F. Pappenberger, D. P. Lettenmeier, R. Buizza, and J. C. Schaake. Application of a medium-range global hydrologic probabilistic forecast scheme to the Ohio River basin. *Wea. Forecasting*, 26: 425–446, 2011.
- M. Vrac and P. Friederichs. Multivariate – inter-variable, spatial and temporal – bias correction. *J. Climate*, 28:218–237, 2015.
- D. S. Wilks. *Statistical Methods in the Atmospheric Sciences*. Academic Press, Amsterdam, 3rd edition, 2011.

- D. S. Wilks. Multivariate ensemble Model Output Statistics using empirical copulas. *Quart. J. Roy. Meteor. Soc.*, 2014. doi: 10.1002/qj.2414. in press.
- R. A. Yuen, T. Gneiting, T. L. Thorarinsdottir, and C. Fraley. *ensembleMOS: Ensemble Model Output Statistics*, 2013. URL <http://CRAN.R-project.org/package=ensembleMOS>. R package version 0.7.